BIOBUSINESS BRIEFS

⮞ REGULATORY WATCH

From big data to smart data: FDA's INFORMED initiative

Recent advances in our understanding of disease mechanisms have led to the development of new drugs that are enabling precision medicine. For example, the co-development of kinase inhibitors that target 'driver mutations' in metastatic non-small-cell lung cancer (NSCLC) with companion diagnostics has led to substantial improvements in the treatment of some patients. However, growing evidence suggests that most patients with metastatic NSCLC and other advanced cancers may not have tumours with single driver mutations. Furthermore, the generation of clinical evidence in genomically diverse and geographically dispersed groups of patients using traditional trial designs and multiple competing therapies is becoming more costly and challenging.

Strategies aimed at creating new efficiencies in clinical evidence generation and extending the benefits of precision medicine to larger groups of patients are driving a transformation from a reductionist approach to drug development (for example, a single drug targeting a driver mutation and traditional clinical trials) to a holistic approach (for example, combination therapies targeting complex multiomic signatures and real-world evidence). This transition is largely fuelled by the rapid expansion in the four dimensions of biomedical big data, which has created a need for greater organizational and technical capabilities (FIG. 1). Appropriate management and analysis of such data requires specialized tools and expertise in health information technology, data science and high-performance computing. For example, efforts to generate clinical evidence using real-world data are being limited by challenges such as capturing clinically relevant variables from vast volumes of unstructured content (such as physician notes) in electronic health records and organizing various structured data elements that are primarily designed to support billing rather than clinical research. So, new standards and quality-control mechanisms are needed to ensure the validity of the design and analysis of studies based on electronic health records.

As a point of convergence for biomedical data intended to support medical product approval decisions, the FDA has taken several steps in recent years to maximize its capabilities for incorporating big data into regulatory decision making. Precision FDA, a research and development portal for validating bioinformatics approaches for processing next-generation sequencing data, and the Sentinel Initiative, a post-market safety surveillance programme based on data from health insurers, are among the examples (see Further information).

Recently, the FDA's Office of Hematology and Oncology Products, in collaboration with the US Department of Health and Human Services' Innovation, Design, Entrepreneurship and Action (IDEA) laboratory, launched the Information Exchange and Data Transformation (INFORMED) initiative to build technical and organizational infrastructure for big data analytics (see Further information and Supplementary information S1 (table)). As a multi-disciplinary effort that draws on the expertise of a range of clinicians, data scientists and entrepreneurs-in-residence, INFORMED has created a new model for collaborative oncology regulatory science research. The activities under INFORMED focus on increasing capabilities for a broad range of data sets, from clinical trials to electronic health records and biosensor technologies. The effort includes a growing portfolio of projects and research collaborations aimed at improving interoperability of disparate information systems, supporting digitization of biomedical content still residing in analogue formats (such as paper and PDF files) and testing new technical frameworks such as blockchain to facilitate secure exchange of health data at scale (*Nat. Rev. Drug Discov.* **15**, 670–671; 2016). It is our hope that by contributing to the advancement of the field of oncology through robust and collaborative regulatory science, we can support transformations that are tuned to the changing needs of cancer patients, together with innovators in the community.

Sean Khozin, Geoffrey Kim and Richard Pazdur are at the US Food and Drug Administration, Office of Hematology and Oncology Products, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993, USA.

Correspondence to S.K.
Sean.Khozin@fda.hhs.gov
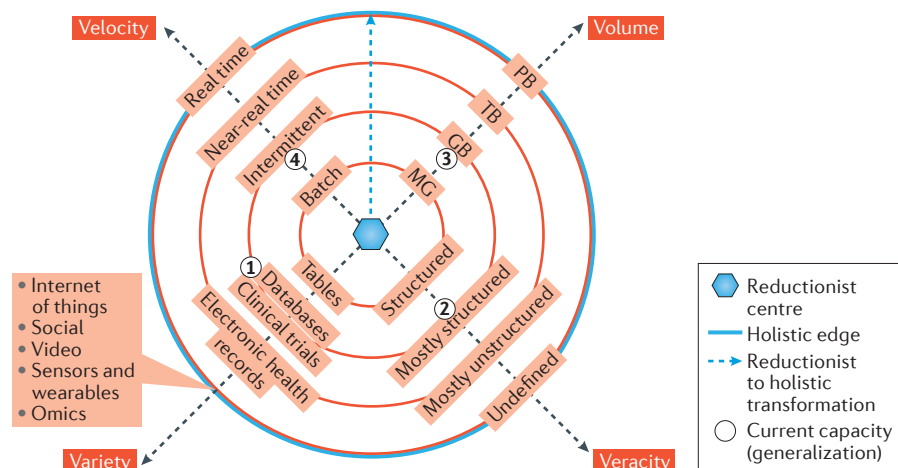
Figure 1 | **Conceptual map of technical and organizational capacity for biomedical big data.** Big data can be defined as having four dimensions: volume (data size), variety (data type), veracity (data noise and uncertainty) and velocity (data flow and processing). Currently, FDA approval decisions are generally based on data of limited variety, mainly from clinical trials and preclinical studies (**1**) that are mostly structured (**2**), in data sets usually no more than a few gigabytes in size (**3**), that are processed intermittently as part of regulatory submissions (**4**). The expansion of big data in the four dimensions (grey lines) calls for increasing organizational and technical capacity. This could transform big data into smart data by enabling a holistic approach to personalization of therapies that takes patient, disease and environmental characteristics into account.

FURTHER INFORMATION
Information Exchange and Data Transformation (INFORMED) initiative: https://www.hhs.gov/idealab/projects-item/information-exchange-and-data-transformation-informed-initiative
Precision FDA: https://precision.fda.gov/
Sentinel Initiative: http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF